

ELDERTON ON CURVES AND CORRELATION

Frequency Curves and Correlation. By W. Palin Elderton. Second edition. London, C. & E. Layton, 1927. viii + 239 pp. Price 15 shillings net + 6d postage.

In such a rapidly developing subject as statistics for an early book to celebrate its twenty-first birthday with an enlarged second edition is phenomenal. The substantial value of the first edition of this book has been enhanced by addition of new material, rearrangement of the old, and a more attractive appearance. The formidable-looking "Key to the Actuarial Terms and Symbols Used" which for the non-actuarial reader stood as a *cave canem* at the beginning of the book has been transferred to the end, and other such rearrangements have been made, rendering the book more readable by avoiding the diversion of attention from the context. There are 67 more pages than in the original edition. This increase is due to the growth in number of Pearson's types of curves from seven to twelve, to the addition of a chapter on partial correlation of the simplest type, and to a somewhat fuller treatment of many topics.

The chief distinctive feature of the book is still the full discussion of the Pearson frequency curves. Now it is true that this system of curves is strictly empirical and does not offer the possibilities, suggested by seminvariants, of incorporating into the curve-fitting process ideas of a general sort regarding properties of compound distributions. Nevertheless they are convenient. Moreover, the modern development of statistics has been so strongly influenced by Karl Pearson that anyone wishing to make his way through statistical literature needs a knowledge of these frequency curves for the same reason that one needs a knowledge of the classics. Such a knowledge may be gained by the abridged reading of the book suggested in an appendix. A considerable part of the book, given to the details of calculation, may be omitted by the casual reader but is indispensable to the curve-fitter.

In a chapter discussing the curves of Edgeworth and the Scandinavians the author concludes on the basis of goodness of fit in examples, and on account of the difficulties connected with series converging, if at all, at uncertain speed, that these systems lack the practical advantages of Pearson's.

The chapter on correlation is radically changed by addition of new material, some of the old material being removed to an appendix. The increased space on correlation is in accord with the increased importance of the subject. The Hardy summation method, which ought to be better known, is applied to the calculation of correlations as well as of univariate moments.

The treatment of correlation of characters not quantitatively measurable presents clearly the chief work of Pearson on the subject. However there ought to be a warning against overworking these methods. In several cases which have come to the reviewer's attention biserial and tetrachoric correlations have been computed with vast labor when all that was wanted was to decide whether there was any correlation at all. These functions, with

their dubious assumption of normal correlation and their large and hazy probable errors, are far less efficient for the purpose than the simpler method of contingency, which yields an immediate and correct interpretation in terms of probability. To take the classic example used in the text, the calculation of an exact value for the correlation between vaccination and recovery from smallpox might well be postponed until, by a much easier and more dependable process, it has been shown that this correlation is significantly greater than zero. So much use of correlation is heuristic that the point is important.

Inverse probability lurks in the book without ever coming into full view. The revolutionary work of R. A. Fisher has found no place in the new edition, so that Chapter X, on probable errors, and Chapter XII, on the correlation ratio and contingency, are partially obsolete. For example, page 187 is directly contradicted by Fisher's conclusion in *Metron* (vol. 5 (1925), No. 3, p. 96). Still more serious from the standpoint of this book are the modifications in the Pearson curve-fitting theory which Fisher has shown to be necessary.

The argument used on page 181 in deriving the formula for the correlation between sampling deviations in class frequencies is unconvincing; for why, in case the number found in a class falls short of expectation, must the corresponding excess in other classes be distributed in exact proportion to their sizes? This can certainly not be expected to happen in all or even in many cases; only by some process of averaging not fully defined can the idea be used. The correlation formula in question may however be reached in a purely algebraic manner by generalizing to n dimensions the ordinary derivation of the standard deviation of a Bernoulli distribution.

On page 48 there is an allusion to some formulas not readily found, and on page 43 it is not clear why, when the criterion is infinite, a root of the quadratic must be infinite rather than zero. It is difficult to see the use or derivation of the functional equation on page 153, and it may possibly mislead the incautious. However the book is on the whole excellent. It supplies a considerable body of valuable information in compact form, and will probably always be the standard book of reference on the Pearson curves.

HAROLD HOTELLING